ENDOGENEOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT

The present invention relates to an endogenous nucleic acid fragment of the retroviral type, integrated into the DNA of the human genome.

Retroviruses are RNA viruses which replicate through a process termed reverse transcription, mediated by an RNA-dependent DNA polymerase named reverse transcriptase (RT), which is encoded by the *pol* gene. The retroviral RNA also comprises at least two additional genes, which are the *gag* and *env* genes. The *gag* gene encodes the proteins of the backbone, i.e. the matrix, the capsid and the nucleocapsid. The *env* gene encodes the envelope proteins. The transcription is regulated by promoter regions located in the LTRs (Long Terminal Repeat) which border the 5'- and 3'-terminal ends of the retroviral genome.

In the course of evolution, humans or their ancestors have integrated material of retroviral origin into their genome subsequent to an infection. Specifically, when a cell is infected, the reverse transcriptase makes a DNA copy of the retroviral RNA, and this DNA copy may then possibly integrate into the human genome. Retroviruses can infect germinal cells and thus be transmitted to future generations by vertical Mendelien transmission. They are then referred to as endogenous retroviruses which are present in the form of proviral DNA integrated into the genome of all human cells. Most endogenous retroviruses are silent or defective. However, some of them have been able to conserve all or part of their initial properties and may be activated under specific conditions. The expression of endogenous retroviruses can range from the transcription of viral genes to the production of viral particles.

These endogenous retroviruses may be associated directly or indirectly with the development of certain pathological conditions.

Endogenous retroviral structures may be in a complete LTR-gag-pol-env-LTR form or in truncated forms.

Thus, in a previous patent application (PCT/FR98/01442), the applicant screened a cDNA library using a Ppol-MSRV probe (SEQ ID NO. 18) and detected overlapping clones which allowed it to reconstruct a putative genomic RNA of 7582 nucleotides. This genomic RNA has an R-U5-gag-pol-env-U3-R structure. A "blastn" interrogation over several databases using the reconstructed genome made it possible to show that there is a considerable amount of related genomic (DNA) sequences in the human genome, which are found on several chromosomes. Thus, the applicant demonstrated the existence of partial structures of the retroviral type in the human genome and envisaged their potential role in the development of autoimmune diseases, in unsuccessful pregnancy or pathological conditions of pregnancy.

Autoimmune diseases which may be mentioned by way of example are multiple sclerosis, rhumatoid arthritis, lupus erythematosus disseminatus, insulin-dependent diabetes and/or pathologies which are associated with them.

The isolation and sequencing of overlapping cDNA fragments and the identification of genomic (DNA) clones corresponding to the isolated DNA clones, described in the applicant's abovementioned PCT patent application, are incorporated herein by way of reference.

Isolation and sequencing of overlapping cDNA fragments:

The information regarding the organization of the novel family of endogenous retroviruses named, by the applicant, HERV-W was obtained by testing a placenta cDNA library (Clontech cat#HL5014a) with the Ppol-MSRV (SEQ ID NO. 18) and Penv-C15 (SEQ ID NO. 19) probes and then carrying out a "gene walking" technique using the novel sequences obtained. The experiments were carried out with reference to the recommendations of the supplier of the library. PCR amplifications on DNA were also used in order to understand this organization.

The following clones were selected and sequenced:

- Clone cl.6A2 (SEQ ID NO. 20): 5' untranslated region of HERV-W and a portion of gag.
- Clone cl.6A1 (SEQ ID NO. 21): gag and a portion of pol.
- Clone cl.7A16 (SEQ ID NO. 22): 3' region of pol.
- Clone cl.Pi22 (SEQ ID NO. 23): 3' region of pol and start of env.
- Clone cl.24.4 (SEQ ID NO. 24): spliced RNA comprising a portion of the 5' untranslated region of HERV-W, the end of pol and the 5' region of env.
- Clone cl.C4C5 (SEQ ID NO. 25): end of env and 3' untranslated region of HERV-W.
- Clone cl.PH74 (SEQ ID NO. 26): subgenomic RNA: 5' untranslated region of HERV-W, end of pol, env, and 3' untranslated region of HERV-W.
- Clone cl.PH7 (SEQ ID NO. 27): multispliced RNA: 5' untranslated region of HERV-W, end of env and 3' untranslated region of HERV-W.
- Clone cl.Pi5T (SEQ ID NO. 28): partial pol gene and U3-R region.
- Clone cl.44.4 (SEQ ID NO. 29): R-U5 region, gag gene and partial pol gene.

A total sequence model for HERV-W was produced with the aid of these clones, by carrying out sequence alignments. The spliced RNAs were revealed and also the potential splice donor and acceptor sites. The LTR, gag, pol and env entities were defined by studying similarity with existing retroviruses.

The putative genetic organization of HERV-W in the RNA form is as follows (SEQ ID NO. 30):

gene 1..7582.

Location of the clones on the reconstructed genomic RNA sequence:
cl.6A2 (1321 bp) 1-1325;
cl.PH74 (535+2229= 2764 bp) 72-606 and 5353-7582;
cl.24.4 (491+1457= 1948 bp); 115-606 and 5353-6810;
cl.44.4 (2372 bp) 115-2496;
cl.PH7 (369+297= 666 bp) 237-606 and 7017-7313; cl.6A1 (2938 bp) 586-3559;
cl.Pi5T (2785+566= 3351 bp) 2747-5557 and 7017-7582;
cl.7A16 (1422 bp) 2908-4337;
cl.Pi22 (317+1689= 2006 bp) 3957-4273 and 4476-6168;
cl.C4C5 (1116 bp) 6467-7582

5'LTR 1..120
/note="R of 5'LTR (5' end uncertain [sic]"
121..575
/note="U5 of 5'LTR"
misc. 579..596
/note="PBS, primer binding site, for tRNA-W"
misc. 606
/note="splice junction (splice donor site ATCCAAAGTG-GTGAGTAATA and splice acceptor site CTTTTTTCAG-ATGGGAAACG, clone RG083M05, GenBank accession AC000064)"
misc. 5353
/note="splice acceptor site for ORF1 (env)"
misc. 5560

/note="splice donor site"
ORF 5581..7194
/note="ORF1 env 538 AA"
/product-="envelope"
misc. 7017
/note="splice acceptor site for ORF2 and ORF3"
ORF 7039..7194
/note="ORF2 52 AA"
ORF 7112..7255
/note="ORF3 48 AA"
misc. 7244..7254
/note="PPT, polypurine tract"
3'LTR 7256..7582
/note="U3-R of 3' LTR (U3-R junction undetermined)
misc. 7563..7569
polyadenylation signal

Identification of genomic (DNA) clones corresponding to the isolated DNA clones:

A "blastn" interrogation over several databases, using the reconstructed genome, showed that there is a considerable amount of related sequences in the human genome. Approximately 400 sequences were identified in GenBank and more than 200 sequences in the EST bank, most of them in the antisense orientation. The 4 most significant sequences in terms of size and similarity are the sequences of the following genomic (DNA) clones:

Human clone RG083M05 (gb AC000064), the chromosomal location of which is 7q21-7q22,

Human clone BAC378 (gb U85196, gb AE000660) corresponding to the alpha/delta locus of the T-cell receptor, located at 14q11-12,

Human cosmid Q11M15 (gb AF045450) corresponding to region 21q22.3 of chromosome 21,

Cosmid U134E6 (embl Z83850) on chromosome Xq22.

The location of the aligned regions for each of the clones is indicated and the chromosome to which they belong is indicated between square brackets (Figure 6 [sic]). The percentage similarity (without the large deletions) between the 4 sequences and the reconstructed genomic RNA is indicated, and also the presence of repeat sequences at each end of the genome and the size of the longest open reading frames (ORFs). Repeat sequences were found at the ends of 3 of these clones. The reconstructed sequence is entirely contained within clone RG083M05 (9.6 Kb) and exhibits 96% similarity. However, clone RG083M05 has a 2 Kb insertion located immediately downstream of the 5′ untranslated region (5′ UTR). This insertion is also found in two other genomic clones which have a 2.3 Kb deletion immediately upstream of the 3′ untranslated region (3′ UTR). No clone contained the three functional gag, pol and env open reading frames (ORFs). Clone RG083M05 shows a 538 amino acid (AA) ORF corresponding to a whole envelope. Cosmid Q11M15 contains two major contiguous ORFs of 413 AA (frame 0) and 305 AA (frame +1) corresponding to a truncated pol polyprotein.

An endogenous nucleic acid fragment has now been found and isolated, which is integrated into the DNA of the human genome and which comprises or consists of at least one portion of the *gag* gene of an endogenous retrovirus associated with an autoimmune disease, or with unsuccessful pregnancy or pathological conditions of pregnancy, this portion at least encoding, directly or indirectly, an expression product. Of course, the invention also comprises the sequence complementary to said fragment.

Advantageously, the fragment defined above also satisfies at least any one of the following characteristics:

It comprises, or consists of, said whole *gag* gene;

Said portion of the fragment at least encodes the matrix and the capsid;

It comprises, or consists of, SEQ ID NO. 1, SEQ ID NO. 2, SEQ ID NO. 3 or the sequence complementary to any one of said sequences;

It is located on at least one of human chromosomes 1, 3, 6, 7 and 16, it is preferably located on at least chromosome 3;

The product of expression of said portion is messenger RNA;

The product of expression of said portion is immunologically recognized by antibodies present in a biological sample from a patient suffering from an autoimmune disease, such as multiple sclerosis; preferably, the biological fluid is chosen from serum, plasma, synovial fluid and urine.

Another subject of the invention is an endogenous transcription product which is in isolated form and which can be obtained by transcription of at least said portion of the *gag* gene of a fragment of the invention.

The invention also relates to a method for detecting endogenous nucleotide sequences belonging to a fragment of the invention, comprising the following steps:

a prior step of extraction of the cellular DNA from a tissue or biological fluid is carried out, and then at

least one cycle of amplification of the cellular DNA is carried out, for instance by PCR, using primers in particular chosen from SEQ ID NO. 4 to SEQ ID NO. 9 and SEQ ID NO. 12 to SEQ ID NO. 17,

the cellular DNA present in the sample is brought into contact with a given probe which is capable of hybridizing with a fragment as defined above and of forming a hybridization complex, said probe comprising at least 15 contiguous nucleotides, preferably 17 and advantageously 19 contiguous nucleotides, of SEQ ID NO. 3, or consisting of SEQ ID NO. 3, under suitable conditions for the hybridization, in particular under conditions of high stringency, and

the hybridization complexes formed are detected by any suitable means.

Advantageously, the probe is labeled with a tracer, such as for example a radioactive tracer or an enzyme.

The invention also relates to a method for detecting endogenous nucleotide sequences belonging to a fragment of the invention, comprising the following steps:

a prior step of extraction of the cellular DNA from a tissue or biological fluid is carried out, and then at least one cycle of amplification of the cellular DNA is carried out, for instance by PCR, using primers in particular chosen from SEQ ID NO. 4 to SEQ ID NO. 9 and SEQ ID NO. 12 to SEQ ID NO. 17,

a step of in vitro transcription/translation of the amplified product is carried out, and

the product derived from the transcription/translation step is reacted with a serum or plasma from a patient with an autoimmune disease.

The invention also relates to a method for studying and/or monitoring T-cell proliferation in vitro, according to which the T cells from a patient are brought into contact with either transcription/translation products (SEQ ID NO. 31), as obtained according to the method above, or synthetic peptides derived from or belonging to SEQ ID NO. 31.

Another subject of the invention is a method for the in situ molecular labeling of chromosomes isolated from patients, in which a probe labeled with any suitable tracer, and comprising all or part of SEQ ID NO. 3, is used.

The invention also relates to:

a recombinant protein obtained using an expression cassette in a bacterial host, characterized in that its protein sequence consists of SEQ ID NO. 31; the bacterial host is in particular *E. coli*;

a reagent for detecting an autoimmune disease or monitoring pregnancy, comprising at least one fragment or one protein of the invention;

the use of a fragment or of a protein of the invention for detecting, in a biological sample, susceptibility to an autoimmune disease, or monitoring pregnancy; in particular, the autoimmune disease is multiple sclerosis.

Before setting out the present invention in greater detail, the definition of certain terms employed in the description and claims is given.

The expression "expression product" means any product derived from the retroviral DNA integrated into the

human genome, including the transcription products (messenger RNA) and the products derived from the translation of the messenger RNA obtained. In the latter case, and by way of example, the product may be a peptide or a protein which is functional or functionalizable, i.e. which can become functional.

The expression "portion encoding, directly or indirectly, an expression product" is intended to mean a portion which, by itself, comprises at least all or part of an open reading frame from which it is possible to deduce an amino acid sequence, and the coding capacity of which can be induced by elements such as, for example, those which may have promoter activity. This definition includes the variability which may be found in the coding nucleic acid sequence, provided that the above conditions are respected.

Example 1: Location of the *gag* gene of the HERV-W family on human chromosomes using the Southern blot technique

In order to locate the *gag* gene of the HERV-W family, a probe corresponding to this gene from MSRV was hybridized on a nylon membrane (Hybond® N+, Amersham) containing 5 µg of DNA from 24 somatic cell hybrids [human x rodents] (isolated human genomic DNA: 22 autosomal chromosomes and 2 sex chromosomes) and 3 control DNAs (human, mouse and hamster), digested with the EcoRI restriction enzyme.

The following probe is used: Pgag-C12 identified by SEQ ID NO. 3 corresponding to the coding region (of 1056 bp) of the clone MSRV gag C12.

1.1- Production of clone 2, C12, containing, in the 3' region, a portion homologous to the *pol* gene, corresponding to the protease gene, and a portion

homologous to the *gag* gene, corresponding to the nucleocapsid, and a 5' coding region, corresponding to the *gag* gene, more specifically the matrix and capsid of MSRV-1.

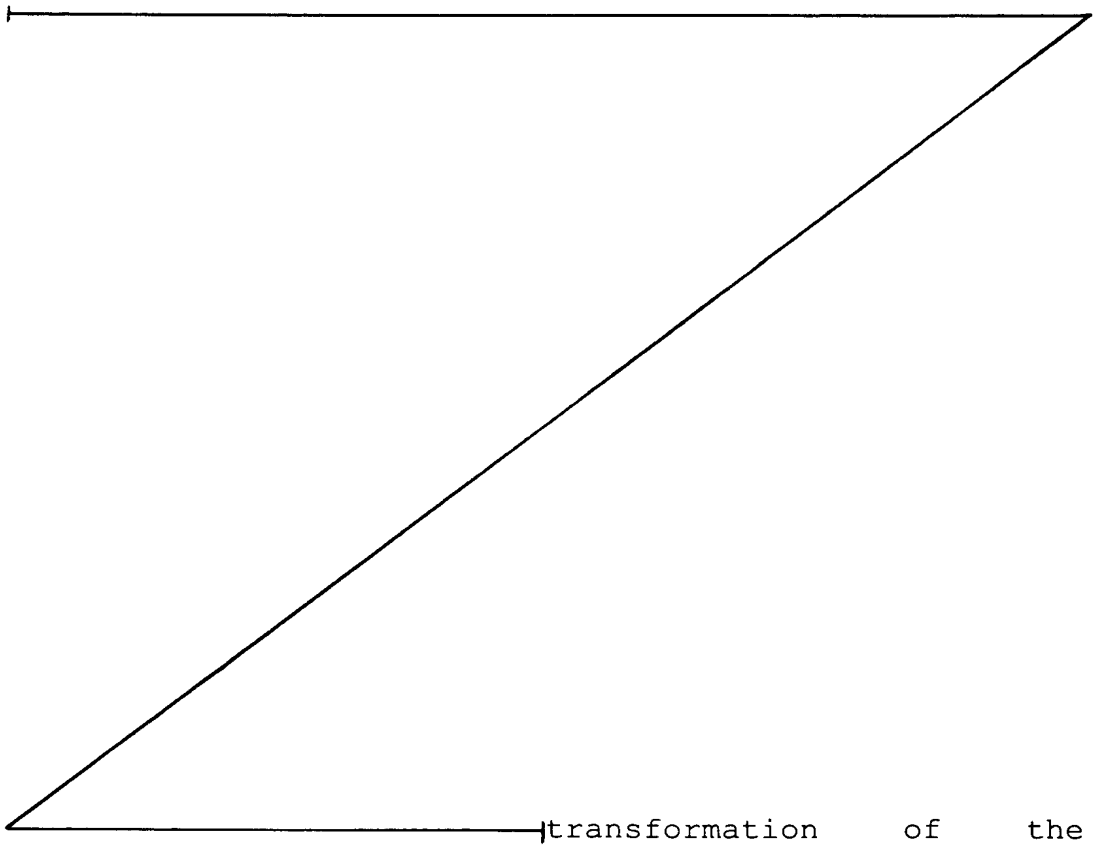
A PCR amplification was carried out on total RNA extracted from 100 µl of plasma from a patient suffering from MS. A water control, treated under the same conditions, was used as a negative control. The cDNA synthesis was carried out with 300 pmol of a random primer (Gibco-BRL, France) and the "Expand RT" reverse transcriptase (Boehringer Mannheim, France), according to the conditions recommended by the company. A PCR (polymerase chain reaction) amplification was carried out with the *Taq* polymerase enzyme (Perkin Elmer, France) using 10 µl of cDNA under the following conditions: 94°C 2 min, 55°C 1 min and 72°C 2 min, then 94°C 1 min, 55°C 1 min and 72°C 2 min for 30 cycles and 72°C for 7 min, with a final reaction volume of 50 µl.

The primers used for the PCR amplification are as follows:

- 5' primer, identified by SEQ ID NO. 4
5' CGG ACA TCC AAA GTG ATG GGA AAC G 3';
- 3' primer, identified by SEQ ID NO. 5
5' GGA CAG GAA AGT AAG ACT GAG AAG GC 3'

A second "nested" PCR amplification was carried out with 5' and 3' primers located inside the region already amplified. This second PCR was carried out under the same experiment conditions as those used in the first PCR, using 10 µl of the amplification product derived from the first PCR.

The primers used for the nested PCR amplification are as follows:



transformation of the ligation in *E. coli* bacteria, the ligation mixture was plated out. At the end of the procedure, the white colonies of recombinant bacteria were picked in order to be cultured and to allow the extraction of the incorporated plasmids according to the "DNA minipreparation" procedure (J. Sambrook, E.F. Fritsch and T. Maniatis, Molecular Cloning, a laboratory manual, Cold Spring Harbour Laboratory Press, 1989). The plasmid preparation from each recombinant colony was cleaved with the Eco RI restriction enzyme and analyzed on agarose gel. The plasmids possessing an insert which was detected under UV light after staining the gel with ethidium bromide were selected in order to sequence the insert after hybridization with a primer complementary to the T7 promoter present on the cloning plasmid from the TA Cloning Kit®. The reaction prior to the sequencing was then carried out according to the method recommended for using the "Prism® Ready Reaction

- 5′ primer, identified by SEQ ID NO. 6
5′ CCT AGA ACG TAT TCT GGA GAA TTG GG 3′;
- 3′ primer, identified by SEQ ID NO. 7
5′ TGG CTC TCA ATG GTC AAA CAT ACC CG 3′

A 1511 bp amplification product was obtained from the RNA extracted from the MS patient plasma. The corresponding fragment was not observed for the water control. This amplification product was cloned in the following way.

The amplified DNA was inserted into a plasmid using the TA Cloning Kit®. The 2 µl of DNA solution were mixed with 5 µl of sterile distilled water, 1 µl of a 10X ligation buffer, 2 µl of PCR® vector (25 ng/ml) and 1 µl of T4 DNA ligase. This mixture was incubated overnight at 14°C. The following steps were carried out in accordance with the instructions of the TA Cloning® kit (Invitrogen). After transformation of the ligation in *E. coli* bacteria, the ligation mixture was plated out. At the end of the procedure, the white colonies of recombinant bacteria were picked in order to be cultured and to allow the extraction of the incorporated plasmids according to the "DNA minipreparation" procedure (J. Sambrook, E.F. Fritsch and T. Maniatis, Molecular Cloning, a laboratory manual, Cold Spring Harbour Laboratory Press, 1989). The plasmid preparation from each recombinant colony was cleaved with the Eco RI restriction enzyme and analyzed on agarose gel. The plasmids possessing an insert which was detected under UV light after staining the gel with ethidium bromide were selected in order to sequence the insert after hybridization with a primer complementary to the T7 promoter present on the cloning plasmid from the TA Cloning Kit®. The reaction prior to the sequencing was then carried out according to the method recommended for using the "Prism® Ready Reaction Amplitaq® FS, DyeDeoxy™ Terminator" sequencing kit

(Applied Biosystems, ref. 402119) and the automatic sequencing was carried out on the Applied Biosystems 373 A and 377 machines, according to the manufacturer's instructions.

The clone obtained, named C12, makes it possible to define a 1511 bp region which has an open reading frame in the N-terminal region of 1056 bp (SEQ ID NO. 3) encoding 359 amino acids (SEQ ID NO. 31) corresponding to the matrix and capsid regions of the *gag* gene.

The nucleotide sequence of C12 is identified by SEQ ID NO. 1. It is represented in Figure 2 with the potential amino acid reading frames.

1.2- Production of the MSRV gag c12 probe

The probe was obtained after PCR amplification, using the pCR™ vector plasmid (TA Cloning® kit, Invitrogen) containing the insert of the clone: MSRV gag c12, with the *Taq* polymerase (Perkin Elmer, France) under the following conditions: 94°C 1 min, 55°C 1 min and 72°C 2 min for 35 cycles and 72°C for 7 min, with a final reaction volume of 100 µl.

The primers used for the PCR amplification are as follows:
- 5′ primer, identified by SEQ ID NO. 12
5′-CTA GAA CGT ATT CTG GAG AAT TGG GA-3′
- 3′ primer, identified by SEQ ID NO. 13
5′-CCT AAG GCA GAC TTT TGA AG-3′.

A 1056 bp amplification product was obtained for MSRV gag c12.

After PCR amplification, the fragment was analyzed in 1% agarose gel. The fragment detected under UV light, after staining the gel with ethidium bromide, was cut

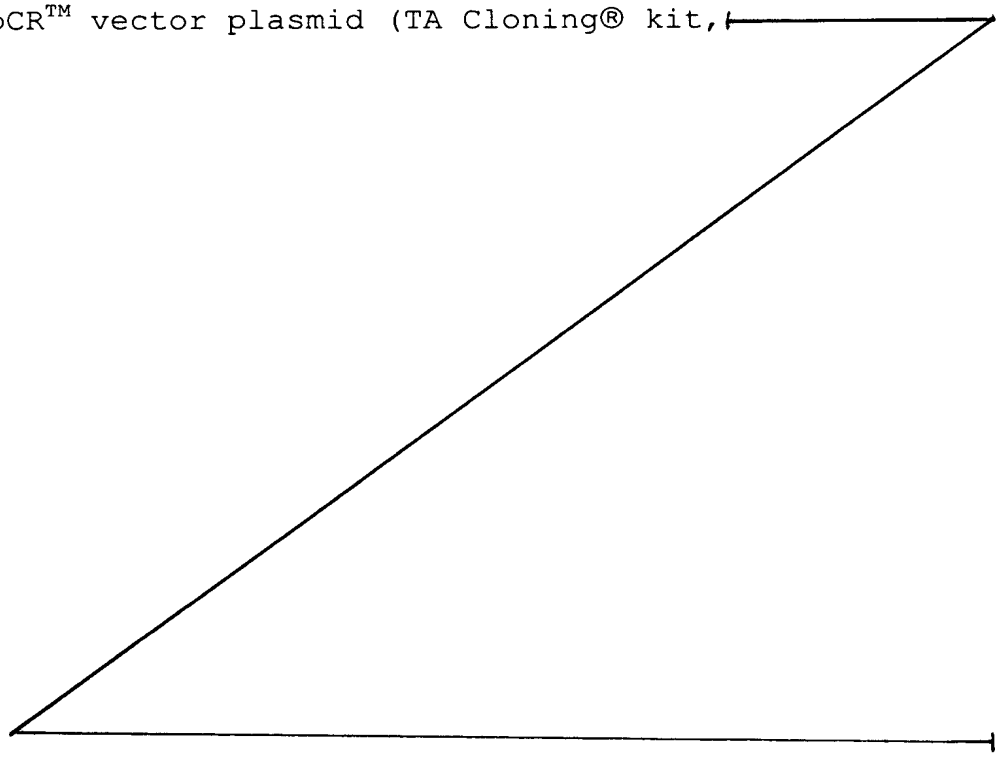
Amplitaq® FS, DyeDeoxy™ Terminator" sequencing kit (Applied Biosystems, ref. 402119) and the automatic sequencing was carried out on the Applied Biosystems 373 A and 377 machines, according to the manufacturer's instructions.

The clone obtained, named Cl2, makes it possible to define a 1511 bp region which has an open reading frame in the N-terminal region of 1089 bp (fragment 434-1521 of SEQ ID NO. 2) encoding 363 amino acids (SEQ ID NO. 31) corresponding to the matrix and capsid regions of the *gag* gene.

The nucleotide sequence of Cl2 is identified by SEQ ID NO. 1. It is represented in Figure 2 with the potential amino acid reading frames.

1.2- Production of the MSRV gag cl2 probe

The probe was obtained after PCR amplification, using the pCR™ vector plasmid (TA Cloning® kit,



out and labeled with [α-P^{32}] using random primers (Gibco-BRL, France) in accordance with instructions of the "Ready-to-go DNA labeling" kit (Pharmacia Biotech). The unincorporated nucleotides were removed with a G-50 Quick Spin column (Boehringer, Mannheim).

1.3- Southern blot

The hybridization conditions are as follows:

After prehybridization for 4 hours (in 5X SSC, 1X Denhardt's, 0.1% SDS, 50% formamide, 20 mM Tris-HCl, pH = 7.5, and 0.1 mg/ml of herring sperm DNA), the nylon membrane containing the human chromosomes was hybridized (in 5X SSC, 1X Denhardt's, 0.1% SDS, 50% formamide, 20 mM Tris-HCl, pH = 7.5, 0.1 mg/ml of herring sperm DNA and 5% dextran sulfate) for 18 hours at 42°C with the ^{32}P-labeled 1056 bp gag cl2 DNA probe (SEQ ID NO. 3). After hybridization, the membrane (The BIOS Monochromosomal Somatic Cell Hybrid blot, from Quantum Bioprobe) hybridized with the gag probe was washed twice in 2X SSC/0.2% SDS solution for 15 min at room temperature, and twice (in 0.2X SSC/0.2% SDS) for 15 min at 45°C. After washing, the membrane was exposed to the X-ray film at -80°C in the presence of an amplifying screen.

The results are given in Table 1 hereinafter.

In this table:

m, which signifies mouse, and h, which signifies hamster, correspond to the recipient cells for the human chromosomal DNA.

The number indicated under each chromosome corresponds to the number of bands encountered.

The total number of copies of the gag gene is 66.

Table 1

Chromo No. Rodent parent	1 m	2 h	3 h	4 h	5 h	6 h	7 h	8 h	9 h	10 h	11 h	12 h	13 h	14 h	15 h	16 m	17 m	18 h	19 h	20 m	21 m	22 h	X h	Y h	Mouse	Ha[illegible] ster
Gag probe	5	0	6	6	5	3	2	3	2	4	3	6	3	1	3	0	3	2	1	0	4	0	4	0	0	0

Example 2: PCR amplification of the *gag* gene of the HERV-W family on each of the isolated human chromosomes; verification of the specificity of the amplifications by Southern blot; "in vitro" transcription/translation (PTT) test using the PCR products, in order to verify the coding capacity and discover which of the human chromosomes have open reading frames for the gag gene of the HERV-W family.

2.1- PCR amplification

In order to amplify the HERV-W *gag* gene, a PCR was carried out on each isolated human chromosome [*NIGMS human/rodent somatic cell hybrid panel #2*. The human monochromosomal NIGMS somatic hybrid mapping panel #2, described by H.L. Drwinga et al. and B.L. Dubois et al., obtained from the Coriell Institute (Camden, NJ)] with the *Taq* polymerase enzyme (Perkin Elmer, France) using: 40 pmol of each primer, 25 mM of each dNTP (Pharmacia), 2.5 mM of $MgCl_2$, 2.5 U of *Taq* polymerase in the standard PCR buffer (Perkin Elmer) and 300 ng of isolated chromosome DNA, in a final volume of 100 µl. The PCR conditions for amplifying the *gag* region are as follows: 3 min at 94°C; then 1 min at 94°C, 1 min at 55°C and 3 min at 72°C for 30 cycles, and 7 min at 72°C.

The primers used for the PCR amplification of the *gag* gene, from an ATG introduced into the HERV-W gag sequence on each isolated human chromosome are as follows:

- 5' primer, identified by SEQ ID NO. 14
5'-TTT GGT AAT ACG ACT CAC TAT AGG GCA GCC ACC ATG GGA AAC GTT CCC CCC GAG-3'.

The primer contains the T7 RNA polymerase promoter sequence, a "spacer", the Kozak sequence (translation

initiation site in eukaryotes) and the 5' gag sequence starting from the HERV-W ATG.
- 3' primer, identified by SEQ ID NO. 15
5'-TTTTTTTTTTTTTTTTTTTCAGGCTGCGCCAGTGTCCAGGAGAC-3'.

The primer contains a poly-A tail (in order to stabilize the transcription of the RNA, represented by 18 T bases), a stop codon (represented by TCA) and the sequence of the MSRV-1 protease gene (G+E+A).

For the amplification of the HERV-W *gag* gene using oligonucleotides defined in the LTR and protease regions of HERV-W, with the *Taq* polymerase enzyme (Perkin Elmer, France), the PCR conditions were as follows: 3 min at 94°C; then 1 min at 94°C, 1 min at 60°C and 2 min at 72°C, 35 cycles; followed by 7 min at 72°C, with 50 ng of each monochromosomal DNA.

The primers used for the PCR amplification of the *gag* gene using the oligonucleotide defined in the HERV-W LTR sequence, on each isolated human chromosome, are as follows:
- 5' primer, identified by SEQ ID NO. 16
5'-TGTCCGCTGTGCTCCTGATC-3'
- 3' primer, identified by SEQ ID NO. 17
5'-TTTTTTTTTTTTTTTTTTTCAGGCTGCGCCAGTGTCCAGGAGAC-3'.

The primer contains a poly-A tail (in order to stabilize transcription of the RNA, represented by 18 T bases), a stop codon (represented by TCA) and the sequence of the MSRV-1 G+E+A protease gene.

The PCR amplifications were carried out in an MJ Research PTC200 Peltier Thermal cycler machine. The PCR products (10 µl of each PCR product) were analyzed in a gel of 1% agarose in 1X TBE (Tris-HCl, borate, EDTA). In order to verify the specificity of the amplification products, 3 µl of each PCR product were analyzed in

agarose gel and then transferred onto a nylon membrane (Hybond® -N^+, Amersham) (Southern blot) using 0.4 N NaOH. The hybridization with the gag cl2 probe (1056 bp) (J. Sambrook et al., 1989) was carried out under the following conditions: after prehybridization for 4 hours (in 5X SSC, 1X Denhardt's, 0.1% SDS, 50% formamide, 20 mM Tris-HCl, pH = 7.5, and 0.1 mg/ml of herring sperm DNA), the nylon membrane was hybridized (in 5X SSC, 1X Denhardt's, 0.1% SDS, 50% formamide, 20 mM Tris-HCl, pH = 7.5, 0.1 mg/ml of herring sperm DNA and 5% dextran sulfate), for 18 hours at 42°C with the ^{32}P-labeled gag DNA probe. The gag PCR products from each isolated human chromosome were washed once, for 15 min at room temperature, in a solution of 2X SSC, 0.2% SDS; twice, for 15 min each wash at 65°C, in a solution of 0.2X SSC, 0.1% SDS; twice, for 15 min each at 65°C, in a solution of 0.1X SSC, 0.1% SDS; and twice, for 30 min each at room temperature, in a solution of 0.1X SSC, 0.1% SDS.

Part of the remaining volume (4 µl) of the PCR amplification products was used for the PTT "in vitro" transcription/translation test (Roest PAM et al., 1993) (Promega, France). The remaining volume was used for the cloning in the pCR® 2.1-TOPO vector (Invitrogen) in accordance with the instructions with the kit, and for the sequencing with the method recommended for using the "PRISM™ Ready Reaction Amplitaq® FS, DyeDeoxy™ Terminator" sequencing kit (Applied Biosystems, ref. 402119), and the automatic sequencing was carried out on Applied Biosystems 373A and 377 machines, according to the manufacturer's instructions.

The portion encoded (SEQ ID NO. 31) by the 2009 bp fragment (SEQ ID NO. 2) was amplified by PCR with the *Pwo* enzyme (5 U/µl) (Boehringer Manneim, France) using 1 µl of the minipreparation of the gag clone DNA (SEQ ID NO. 3) under the following conditions: 95°C 1 min,

60°C 1 min and 72°C 2 min for 25 cycles, with a final reaction volume of 50 µl, using the primers:
- 5′ primer (*Bam* HI) (SEQ ID NO. 8):
5′ ATG GGA AAC GTT CCC CCC GAG 3′ (21 mer), and
- 3′ primer (*Hind* III), identified by SEQ ID NO. 9
5[sic] GGC CTA AGG CAG ACT TTT GAA 3′ (21 mer).

The fragment obtained after PCR was linearized with *Bam* HI and *Hind* III and subcloned into the pET28C and pET21C expression vectors (NOVAGEN) linearized with *Bam* HI and *Hind* III. The DNA of the 1089 bp fragment in the two expression vectors were sequenced according to the method recommended for using the "PRISM™ Ready Reaction Amplitaq® FS, DyeDeoxy™ Terminator" sequencing kit (Applied Biosystems, ref. 402119) and the automatic sequencing was carried out on Applied Biosystems 373A and 377 machines, according to the manufacturer's instructions.

The expression of the nucleotide sequence of the 1089 bp fragment of the gag clone by the pET28C and pET21C expression vectors is identified by SEQ ID NO. 10 and SEQ ID NO. 11, respectively.

2.2- "In vitro" transcription/translation test (PTT, Promega)

This test was carried out in order to pinpoint the human chromosomes which have open reading frames for the *gag* gene of the HERV-W family.

A mixture containing 12.5 µl of TNT® rabbit reticulocyte lysate (Promega), 1 µl of TNT® reaction buffer (Promega), 0.5 µl of TNT® RNA polymerase (Promega), 0.5 µl of a 1 mM mixture of amino acids minus methionine, 2 µl of ^{35}S-methionine (1000 Ci/mmol) at 10 mCi/µl (Amersham), 0.5 µl of RNasin® ribonuclease inhibitor at 40 U/µl, 4 µl of PCR amplification

products (equivalent to 1 µg) from each human chromosome and 4 µl of water, in a reaction volume of 25 µl, [lacuna]. This mixture was incubated at 30°C for 90 min.

The gag proteins corresponding to the products of transcription/translation of the *gag* gene of the HERV-W family from each human chromosome, amplified by PCR, were revealed by 10% polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate (SDS)-PAGE after exposure of the gel to the X-ray film at room temperature in the presence of an amplifying screen.

The results are given in Table 2 hereinafter.

In this table, the number indicated under each chromosome corresponds to the molecular mass (kDa) of the proteins visualized in polyacrylamide gel in the presence of SDS.

Table 2

Chromo No. Rodent parent	Chromo No. Rodent parent	1 m	2 h	3 h	4 h	5 h	6 h	7 h	8 h	9 h	10 h	11 h	12 h	13 h	14 h	15 h	16 m	17 m	18 h	19 h	20 m	21 m	22 h	X h	Y h	Water
PCR gag	gag	28 23 18		45 25 20 17			23 18 17	22									14									

Example 3: Expression of the gag clone in *Escherichia coli*, and reaction with human sera

The coding region SEQ ID NO. 2 was expressed in *Escherichia coli*, and then the products thus expressed were tested against serum from patients suffering from MS, and also against serum from healthy patients.

The constructs pET28c-gag clone (1089 bp) and pET21C-gag clone (1089 bp) synthesize, in the BL21 (DE3) bacterial strain, an N-terminal and C-terminal fusion protein for the pET28C vector, and a C-terminal fusion protein for the pET21C vector, with 6 histidine residues and an apparent molecular mass of approximately 45 kDa, which are revealed by SDS-PAGE polyacrylamide gel electrophoresis (U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature, 1970, 227: 680–685).

The reactivity of the protein with respect to an anti-histidine monoclonal antibody (DIANOVA) was demonstrated using the Western blot technique (H. Towbin et al., Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications, Proc. Natl. Acad. Sci. USA, 1979, 76: 4350-4354).

The recombinant proteins pET28C-gag clone (1089 bp) and pET21C-gag clone (1089 bp) were visualized, by SDS-PAGE, in the insoluble fraction after enzymatic digestion of the bacterial extracts with 50 µl of lysozyme (10 mg/ml) and lysis by ultrasound.

The antigenic properties of the recombinant antigens pET28C-gag clone (1089 bp) and pET21C-gag clone (1089 bp) were tested by Western blot after solubilization of the bacterial pellet with 2% SDS and

50 mM β-mercaptoethanol. After incubation with the sera from patients suffering from multiple sclerosis, the sera from the neurological controls and the blood transfusion center (BTC) control sera, the immunocomplexes were detected using an alkaline phosphatase-coupled anti-human IgG and IgM goat serum.

The results are given in Table 3 hereinafter.

Table 3

Reactivity of sera from patients suffering from multiple sclerosis and controls, with the recombinant gag protein produced in *E. coli*[a]

DISEASE	NUMBER OF INDIVIDUALS TESTED	NUMBER OF POSITIVE INDIVIDUALS
MS	15	6 2(+++), 2(++), 2(+)
NEUROLOGICAL CONTROLS	2	1(+++)
HEALTHY CONTROLS (BTC)	22	1(+/-)

(a) The strips containing 1.5 µg of recombinant gag antigen show reactivity against sera diluted 100-fold. The Western blot interpretation is based on the presence or absence of a gag-specific band on the strips. Positive and negative controls are included in each experiment.

These results show that, under the technical conditions used, approximately 40% of the human multiple sclerosis sera tested react with the recombinant gag protein.